

Efficient algorithms computing distances between Radon measures on \mathbb{R}

Jędrzej Jabłoński¹ and Anna Marciniak-Czochra²

¹ *Institute of Applied Mathematics and Mechanics,
University of Warsaw
Warszawa 02-097, Poland*

² *Institute of Applied Mathematics,
Interdisciplinary Center for Scientific Computing (IWR),
and BIOQUANT Center
University of Heidelberg
Im Neuenheimer Feld 294, 69120 Heidelberg, Germany*

April 15, 2013

Abstract

In this paper numerical methods of computing distances between two Radon measures on \mathbb{R} are discussed. Efficient algorithms for Wasserstein-type metrics are provided. In particular, we propose a novel algorithm to compute the flat metric (bounded Lipschitz distance) with a computational cost $O(n \log n)$. The flat distance has recently proven to be adequate for the Escalator Boxcar Train (EBT) method for solving transport equations with growth terms. Therefore, finding efficient numerical algorithms to compute the flat distance between two measures is important for finding the residual error and validating empirical convergence of different methods.

Keywords: metric spaces, flat metric, Wasserstein distance, Radon measures, optimal transport, linear programming, minimum-cost flow

1 Introduction

Recent years witnessed large developments in the kinetic theory methods applied to mathematical physics and more recently also to mathematical biology. Among important branches of the kinetic theory are optimal transportation problems and related to them Wasserstein metrics, and Monge-Kantorovich metrics [2, 14]. Partial differential equations in metric spaces are being applied to transportation problems [10, 14], gradient flows [2, 13] and structured population models [4, 7, 8, 12].

Output of mathematical modeling can often be described as Radon measures. Comparing the results of the models requires then a definition of distance in the space of measures. The desired properties of such metrics depend on the structure of the considered problem. In most of the cases, the topology of total variation is too strong for applications, and weaker metrics had to be introduced, see [8] for details. Well known 1-Wasserstein metric, on the other hand, is only applicable to processes with mass conservation. To cope with growth in the process, various modifications have been proposed, including flat metric and centralized Wasserstein metric. In the present paper, we additionally introduce a normalized Wasserstein distance. For comparison of different metrics, their interpretation and examples see Table 1. Even though, all those distances can be computed using linear programming (LP), its computational complexity becomes often larger than the complexity of solving the original problem. For example, the equations for which the stability of numerical algorithm is proven in $(W^{1,\infty})^*$ require an efficient algorithm for the flat metric to find the residual error.

Algorithms proposed in this paper are designed to compute efficiently Wasserstein-type distance between two Radon measures in the form of $\sum_{i=1}^n a_i \delta_{x_i}$, where δ is the Dirac delta. The algorithms can be also applied for the case of an arbitrary pair of measures, by approximating those measures by the sum of Dirac deltas. In the process of finding numerical solutions to partial differential equations, the distance between a discrete and absolutely continuous measures is needed to evaluate the quality of the initial condition approximation, while the distance between two discrete measure is needed to find the residual error.

The paper is organized as follows. In Section 2 we introduce and compare four Wasserstein-type metrics. Additionally, we introduce a two-argument function, based on the Wasserstein distance, which is not a metric, however, provides a good tool to estimate the flat metric from above. Section 3 is devoted to numerical algorithms proposed to calculate distances between two measures in respect to the considered metrics. We justify the algorithms and provide respective pseudocodes. The novelty of this paper is the algorithm for flat metric, which has been recently proven to be adequate for the Escalator Boxcar Train (EBT) method for solving transport equations [3]. We propose an efficient algorithm which computation cost is $O(n^2)$ and optimize it further to $O(n \log n)$. To judge the efficiency of the algorithms, we compare the times needed to compute the flat distance between two sums of a large number of Dirac deltas randomly distributed over $[-1,1]$.

2 Metrics on the spaces of Radon measures

2.1 1-Wasserstein distance

The framework of Wasserstein metric in the spaces of probability measures has proven to be very useful for the analysis of equations given in a conservative form, such as, for example, transport equation [2], Fokker-Planck [9] or nonlinear diffusion equation [5]. It was originally defined using the notion of optimal transportation, but we focus on its dual

representation.

Definition 2.1. 1-Wasserstein distance between two measures μ and ν is given by

$$W_1(\mu, \nu) = \sup \left\{ \left| \int_{\mathbb{R}} f(x) d(\mu - \nu)(x) \right| : f \in C(\mathbb{R}, \mathbb{R}), \text{Lip}(f) \leq 1 \right\} \quad (1)$$

Observe that the integral $\int_{\mathbb{R}} f(x) d(\mu - \nu)(x)$ is invariant with respect to adding a constant to function f . Consequently, for any $x^* \in \mathbb{R}$ holds

$$W_1(\mu, \nu) = \sup \left\{ \int_{\mathbb{R}} f(x) d(\mu - \nu)(x) : f \in C(\mathbb{R}, \mathbb{R}), f(x^*) = 0, \text{Lip}(f) \leq 1 \right\}.$$

Furthermore, the following two properties hold:

- 1-Wasserstein distance is scale-invariant

$$W_1(\lambda \cdot \mu, \lambda \cdot \nu) = \lambda W_1(\mu, \nu).$$

- 1-Wasserstein distance is translation-invariant

$$W_1(T_x \mu, T_x \nu) = W_1(\mu, \nu).$$

2.2 Normalized Wasserstein metric

If $\mu(\mathbb{R}) \neq \nu(\mathbb{R})$, then $W_1(\mu, \nu) = \infty$ by the dual representation definition. It makes this metric useless for processes not preserving the conservation of mass.

In this section we introduce a simple normalization that leads to a definition of translation-invariant metric suitable for non-local models

Definition 2.2. We define normalized 1-Wasserstein distance between two measures μ and ν as

$$\widetilde{W}_1(\mu, \nu) = \min \left(\|\mu\| + \|\nu\|, |\|\mu\| - \|\nu\|| + W_1 \left(\frac{\mu}{\|\mu\|}, \frac{\nu}{\|\nu\|} \right) \right), \quad (2)$$

where $\|\mu\|$ denotes the total variation of the measure μ .

Lemma 2.3. The distance defined by (2) is a metric.

Proof. Let μ, ν and η be Radon measures. Then, it holds

1. $\widetilde{W}_1(\mu, \nu) = 0$ if and only if $\mu = \nu$.

Indeed, either $\|\mu\| + \|\nu\| = 0$ or $|\|\mu\| - \|\nu\|| + W_1 \left(\frac{\mu}{\|\mu\|}, \frac{\nu}{\|\nu\|} \right) = 0$ imply that $\mu = \nu$.

2. $\widetilde{W}_1(\mu, \nu) = \widetilde{W}_1(\nu, \mu),$

3. Since

$$\begin{aligned} \widetilde{W}_1(\mu, \nu) + \widetilde{W}_1(\nu, \eta) &= \min \left(\|\mu\| + \|\nu\|, |\|\mu\| - \|\nu\|| + W_1 \left(\frac{\mu}{\|\mu\|}, \frac{\nu}{\|\nu\|} \right) \right) \\ &\quad + \min \left(\|\eta\| + \|\nu\|, |\|\eta\| - \|\nu\|| + W_1 \left(\frac{\eta}{\|\eta\|}, \frac{\nu}{\|\nu\|} \right) \right), \end{aligned}$$

to show the triangle inequality, we consider four possibilities

$$\begin{aligned} \widetilde{W}_1(\mu, \nu) + \widetilde{W}_1(\nu, \eta) &= \|\mu\| + \|\nu\| + \|\eta\| + \|\nu\| \geq \|\mu\| + \|\eta\| \geq \widetilde{W}_1(\mu, \eta), \\ \widetilde{W}_1(\mu, \nu) + \widetilde{W}_1(\nu, \eta) &= |\|\mu\| - \|\nu\|| + W_1 \left(\frac{\mu}{\|\mu\|}, \frac{\nu}{\|\nu\|} \right) \\ &\quad + |\|\eta\| - \|\nu\|| + W_1 \left(\frac{\eta}{\|\eta\|}, \frac{\nu}{\|\nu\|} \right) \geq \widetilde{W}_1(\mu, \eta), \\ \widetilde{W}_1(\mu, \nu) + \widetilde{W}_1(\nu, \eta) &= \|\mu\| + \|\nu\| + |\|\eta\| - \|\nu\|| + W_1 \left(\frac{\eta}{\|\eta\|}, \frac{\nu}{\|\nu\|} \right) \geq \|\mu\| + \|\eta\| \\ &\geq \widetilde{W}_1(\mu, \eta), \\ \widetilde{W}_1(\mu, \nu) + \widetilde{W}_1(\nu, \eta) &= \|\eta\| + \|\nu\| + |\|\mu\| - \|\nu\|| + W_1 \left(\frac{\mu}{\|\mu\|}, \frac{\nu}{\|\nu\|} \right) \geq \|\mu\| + \|\eta\| \\ &\geq \widetilde{W}_1(\mu, \nu). \end{aligned}$$

□

This metric can be easily computed numerically using the algorithm for 1-Wasserstein distance. However, it lacks the scaling property, which holds for the Wasserstein distance, and which is useful for applications. Nonetheless, the following weaker property holds.

Proposition 2.4. *Let μ_k and ν_k be two sequences of Radon measures and $\|\mu_k\| \rightarrow 0$, $\|\nu_k\| \rightarrow 0$ then $\widetilde{W}_1(\mu_k, \nu_k) \rightarrow 0$*

2.2.1 Centralized Wasserstein metric

In this section we present a different modification of the 1-Wasserstein distance, which is scale-invariant.

Definition 2.5. *We define centralized 1-Wasserstein distance between two measures μ and ν as*

$$\widehat{W}_1(\mu, \nu) = \sup \left\{ \left| \int_{\mathbb{R}} f(x) d(\mu - \nu)(x) \right| : f \in C(\mathbb{R}, \mathbb{R}), \text{Lip}(f) \leq 1, f(0) \in [-1, 1] \right\}.$$

The metric was introduced in [8] for analysis of the structured population models in the spaces of non-negative Radon measures, $\mathcal{M}_1^+(\mathbb{R}^+)$, on \mathbb{R}^+ with integrable first moment, i.e.,

$$\mathcal{M}_1(\mathbb{R}^+) := \left\{ \mu \in \mathcal{M}^+(\mathbb{R}^+) \mid \int_{\mathbb{R}^+} |x| d\mu < \infty \right\}.$$

This metric satisfies the scaling property, but on the contrary to Wasserstein metric, it is not invariant with respect to translations. It is therefore only useful for modeling specific phenomena, for example structured population dynamics where mass generation can only occur in one specific point of space.

2.3 Flat metric

Another solution of the problem of comparing two measures of different masses is the flat metric which is defined as follows

Definition 2.6. *Flat distance between two measures μ and ν is given by*

$$F(\mu, \nu) = \sup \left\{ \left| \int_{\mathbb{R}} f(x) d(\mu - \nu)(x) \right| : f \in B_{C(\mathbb{R}, \mathbb{R})}(0, 1), \text{Lip}(f) \leq 1 \right\}.$$

The flat metric, known also as a bounded Lipschitz distance [11], corresponds to the dual norm of $W^{1,\infty}(\mathbb{R})$, since the test functions used in the above definition are dense in $W^{1,\infty}(\mathbb{R})$. This metric satisfies the scaling property and it is translation invariant. It has proven to be useful in analysis of structured population models and in particular, Lipschitz dependence of solutions on the model parameters and initial data [7, 4]. The flat metric has been recently used for the proof of convergence and stability of EBT, which is a numerical algorithm based on particle method, for the transport equation with growth terms [3]. Consequently, an implementation of the EBT method requires an algorithm to compute the flat distance between two measures.

2.4 Upper bound for flat metric

Computing of flat metric exactly is more costly than computing of 1-Wasserstein metric ($O(n \log n)$ vs $O(n)$, see Section 3.2.4). Moreover, often in applications it is sufficient to calculate upper bound of the distance, for example when computing residual error. We propose the following function, which requires only linear computing time.

$$\widehat{F}(\mu, \nu) = | \|\mu\| - \|\nu\| | + \begin{cases} W_1 \left(\mu, \nu \frac{\|\mu\|}{\|\nu\|} \right) & \text{if } \|\mu\| < \|\nu\| \\ W_1 \left(\mu \frac{\|\nu\|}{\|\mu\|}, \nu \right) & \text{if } \|\mu\| \geq \|\nu\| \end{cases}$$

Lemma 2.7. *Let μ, ν be Radon measures on \mathbb{R} . Then, the following estimate holds*

$$F(\mu, \nu) \leq \widehat{F}(\mu, \nu)$$

Proof. Assume that $\|\mu\| < \|\nu\|$. Then, using the estimate of the flat metric by 1-Wasserstein distance for the measures with equal masses provides

$$\begin{aligned}
F(\mu, \nu) &\leq F\left(\mu, \nu \cdot \frac{\|\mu\|}{\|\nu\|}\right) + F\left(\nu \cdot \frac{\|\mu\|}{\|\nu\|}, \nu\right) \\
&\leq W_1\left(\mu, \nu \cdot \frac{\|\mu\|}{\|\nu\|}\right) + F\left(\nu \cdot \frac{\|\mu\|}{\|\nu\|}, \nu\right) \\
&= W_1\left(\mu, \nu \cdot \frac{\|\mu\|}{\|\nu\|}\right) + \sup \left\{ \left| \int_{\mathbb{R}} f(x) \left(\frac{\|\mu\|}{\|\nu\|} - 1 \right) d\nu(x) \right| : f \in B_{C(\mathbb{R}, \mathbb{R})}(0, 1), \text{Lip}(f) \leq 1 \right\} \\
&= W_1\left(\mu, \nu \cdot \frac{\|\mu\|}{\|\nu\|}\right) + \left| \frac{\|\mu\|}{\|\nu\|} - 1 \right| \|\nu\|.
\end{aligned}$$

The last equality results from the representation of Radon distance and positivity of measure ν . \square

The upper bound function is more useful if it can be estimated from above by $c \cdot F(\mu, \nu)$ for some constant c . In a general case, however, such constant does not exist, since by taking $\mu = \delta_0$, $\nu = \delta_x$ and passing $x \rightarrow \infty$ we obtain

$$\begin{aligned}
\widehat{F}(\mu, \nu) &\rightarrow \infty, \\
F(\mu, \nu) &\rightarrow 2.
\end{aligned}$$

Nevertheless, the desired estimate can be shown on a bounded set.

Lemma 2.8. *Let μ, ν be non-negative Radon measures on a compact set K , then the following estimate holds*

$$\widehat{F}(\mu, \nu) \leq c_K F(\mu, \nu)$$

Proof. Let $c_K = \max(1, \frac{1}{2} \text{diam}(K))$ and x_0 be such point that $|k - x_0| \leq c_K$ for any $k \in K$. Then, assuming $\|\mu\| < \|\nu\|$, we obtain

$$\begin{aligned}
\widehat{F}(\mu, \nu) &= | \|\mu\| - \|\nu\| | + W_1 \left(\mu, \nu \frac{\|\mu\|}{\|\nu\|} \right) \\
&= \left| \int d(\mu - \nu) \right| + W_1 \left(\mu, \nu \frac{\|\mu\|}{\|\nu\|} \right) \\
&\leq F(\mu, \nu) + W_1 \left(\mu, \nu \frac{\|\mu\|}{\|\nu\|} \right) \\
&\leq F(\mu, \nu) + W_1(\mu, \nu) + W_1 \left(\nu, \nu \frac{\|\mu\|}{\|\nu\|} \right) \\
&\leq 2 \cdot F(\mu, \nu) + W_1(\mu, \nu) \\
&= 2 \cdot F(\mu, \nu) + \sup \left\{ \left| \int_{\mathbb{R}} f(x) d(\mu - \nu)(x) \right| : f \in C(\mathbb{R}, \mathbb{R}), \text{Lip}(f) \leq 1 \right\} \\
&= 2 \cdot F(\mu, \nu) + c_K \sup \left\{ \left| \int_{\mathbb{R}} f(x) d(\mu - \nu)(x) \right| : f \in C(\mathbb{R}, \mathbb{R}), \text{Lip}(f) \leq \frac{1}{c_K} \right\} \\
&\leq (2 + c_K) F(\mu, \nu)
\end{aligned}$$

□

2.5 Comparison of the presented metrics

In the following Table we compare the introduced metrics and provide examples and interpretations in terms of optimal transport.

Metric	Example: $d(2\delta_x, 3\delta_y)$	Scale- invariance	Translation- invariance	Intuition of $d(\mu, \nu)$	Compute complexity
Wasser- stein	∞	YES	YES	The cost of optimal transportation of distribution μ to a state given by ν , assuming that moving mass m by x requires mx energy.	$O(n)$

Wasser- stein normal- ized	$\min(2 + 3, (3 - 2) + x - y)$	weak	YES	Minimum of the cost of annihilating μ and generating ν ; and of the cost of generating the difference in mass between μ and ν and transporting $\frac{\mu}{\ \nu\ }$ to $\frac{\nu}{\ \nu\ }$, assuming that generating/annihilating mass at any position requires the cost equal to the mass.	$O(n)$
Wasser- stein central- ized	$2 x - y + y $	YES	NO	The cost of generating the difference in mass in point 0 in space added to the cost of transporting $\mu + (\ \nu\ - \ \mu\) \delta_0$ to ν .	$O(n)$
Flat	$1 + 2 \min(2, x - y)$	YES	YES	The cost of optimal transporting AND/OR generating AND/OR annihilating mass to form ν from μ .	$O(n \log n)$
Radon	$2 + 3$	YES	YES	The cost of generating AND/OR annihilating mass to form ν from μ	$O(n)$

Table 1: Comparison of different metrics.

3 Approximations and computational algorithms

In the remainder of this paper, we present algorithms for computing and approximating of the introduced metrics.

3.1 Computing of 1-Wasserstein distance

We start with introducing the algorithm for computation of 1-Wasserstein distance. The algorithm is well-known and straightforward. Nevertheless, we present it here, since the specific approach used in this section will be also applied later in the proof of correctness of more involved algorithms for other distances.

3.1.1 Reduction to the case of discrete measures

Lemma 3.1. *Let μ be an arbitrary Radon measure supported on $[0, 1]$. For every $\varepsilon > 0$ there exist a discrete measure μ_ε such that*

$$W_1(\mu, \mu_\varepsilon) < \varepsilon$$

Proof. Define

$$\mu_{\frac{\mu[0,1]}{n}} = \sum_{i=1}^n \delta_{i/n} \mu \left[\frac{i-1}{n}, \frac{i}{n} \right).$$

Then, we estimate

$$W_1(\mu, \mu_{\frac{\mu[0,1]}{n}}) \leq \sum_{i=1}^n \frac{1}{n} \mu \left[\frac{i-1}{n}, \frac{i}{n} \right) \leq \frac{1}{n} \mu[0, 1],$$

and the right-hand side tends to 0 with $n \rightarrow \infty$. □

Consequently, 1-Wasserstein distance between an arbitrary pair of finite Radon measures can be estimated by the distances between their discrete approximations. From this point on, we assume that

$$\mu - \nu = \sum_{k=1}^n a_k \delta_{x_k}.$$

3.1.2 The algorithm

The formula for 1-Wasserstein distance reads

$$W_1(\mu, \nu) = \sup \left\{ \sum_{k=1}^n a_k f(x_k) : f \in C(\mathbb{R}, \mathbb{R}), f(x_n) = 0, Lip(f) \leq 1 \right\}.$$

Regularity conditions can be represented as linear programming bounds. Hence, computing of $W_1(\mu, \nu)$ is equivalent to finding maximum of

$$\left| \sum_{k=1}^n a_k f_k \right|$$

with the following restrictions

$$\begin{aligned} f_n &= 0, \\ |f_k - f_{k-1}| &\leq |x_k - x_{k-1}|. \end{aligned}$$

Although this problem can be solved by linear programming, a more efficient algorithm can be found.

Define

$$W_1^m(x) = \sup \left\{ \sum_{k=1}^m a_k f_k : f_m = x, |f_k - f_{k-1}| \leq |x_k - x_{k-1}| \right\}.$$

Then, obviously $W_1(\mu, \nu) = W_1^n(0)$. Denote $d_k = x_{k+1} - x_k$, and observe that

$$W_1^1(x) = a_1 x,$$

$$\begin{aligned} W_1^2(x) &= a_2 x + \sup_{f_1 \in [x-d_1, x+d_1]} W_1^1(x) = a_2 x + a_1 x + a_1 \cdot \text{sgn}(a_1) d_1 = \\ &= (a_1 + a_2) x + |a_1| d_1. \end{aligned}$$

It can be shown by induction that

$$W_1^n(x) = \left(\sum_{i=1}^n a_i \right) x + \sum_{i=1}^{n-1} d_i \left| \sum_{j=1}^i a_j \right|.$$

Notice that the value a_n is not used in this formula. It is, however, involved indirectly, as $\sum_{i=1}^n a_i = 0$.

3.1.3 Pseudocode

Input: Non-decreasing table of positions $x \in [0, 1]^n$, table of masses $a \in \mathbb{R}^n$
1-WASSERSTEIN-DISTANCE ($x \in [0, 1]^n$, $a \in \mathbb{R}^n$)

```

distance  $\leftarrow$  0
partialSum  $\leftarrow$  0
for idx  $\leftarrow$  1 to  $n-1$  do
    partialSum  $\leftarrow$  partialSum +  $a_{idx}$ 
    distance  $\leftarrow$  distance +  $(x_{idx+1} - x_{idx}) \cdot |partialSum|$ 
return distance

```

3.1.4 Complexity of the algorithm

It is clear from the pseudocode that the computational complexity of the algorithm is $\Theta(n)$, while memory complexity is $\Theta(1)$.

3.2 Computing of the centralized Wasserstein distance

3.2.1 Reduction to the case of discrete measures

Similarly to the case of 1-Wasserstein distance the following lemma holds.

Lemma 3.2. *Let μ be an arbitrary Radon measure supported on $[0, K]$. For every $\varepsilon > 0$, there exists a discrete measure μ_ε such that*

$$\widehat{W}_1(\mu, \mu_\varepsilon) < \varepsilon.$$

The proof follows from the fact that $\|\mu\| = \|\nu\|$ implies $\widehat{W}_1(\mu, \nu) = W_1(\mu, \nu)$.

3.2.2 The Algorithm

We assume

$$\mu - \nu = \sum_{i=1}^m a_i \delta_{x_i} + a_{m+1} \delta_0 + \sum_{i=m+2}^n a_i \delta_{x_i}.$$

Define

$$\begin{aligned} \underline{W}_1^j(x) &= \sup \left\{ \sum_{k=1}^j a_k f_k : f_j = x, |f_k - f_{k-1}| \leq |x_k - x_{k-1}| \right\}, \\ \overline{W}_1^j(x) &= \sup \left\{ \sum_{k=j}^n a_k f_k : f_j = x, |f_k - f_{k-1}| \leq |x_k - x_{k-1}| \right\}. \end{aligned}$$

As already proven

$$\begin{aligned} \underline{W}_1^{m+1}(x) &= \left(\sum_{i=1}^{m+1} a_i \right) x + \sum_{k=1}^m d_k \left| \sum_{i=1}^k a_i \right|, \\ \overline{W}_1^{m+1}(x) &= \left(\sum_{i=m+1}^n a_i \right) x + \sum_{k=1}^{n-(m+1)} d_{n-k} \left| \sum_{i=n+1-k}^n a_i \right|. \end{aligned}$$

From LP representation of the metric

$$\widehat{W}_1(\mu, \nu) = \sup \left\{ \sum_{k=1}^n a_k f_k : -1 \leq f_{m+1} \leq 1, |f_k - f_{k-1}| \leq |x_k - x_{k-1}| \right\},$$

it can be deduced that

$$\widehat{W}_1(\mu, \nu) = \sup_{x \in [-1, 1]} \left(\underline{W}_1^{m+1}(x) + \overline{W}_1^{m+1}(x) - a_{m+1}x \right),$$

so the distance is given by the formula

$$\widehat{W}_1(\mu, \nu) = \sum_{k=1}^m d_k \left| \sum_{i=1}^k a_i \right| + \sum_{k=1}^{n-(m+1)} d_{n-k} \left| \sum_{i=n+1-k}^n a_i \right| + \left| \sum_{i=1}^n a_i \right|.$$

3.2.3 Pseudocode

Input: Non-decreasing table of positions $x \in [0, 1]^k \times \{0\} \times [0, 1]^{n-k-1}$, table of masses $a \in \mathbb{R}^n$
1-WASSERSTEIN-CENTRALIZED-DISTANCE

```

distance  $\leftarrow$  0
(partialSumFront, partialSumBack)  $\leftarrow$  (0, 0)
(idxFront, idxBack)  $\leftarrow$  (1,  $n$ )
while  $x_{idxFront} < 0$  do
    partialSumFront  $\leftarrow$  partialSumFront +  $a_{idxFront}$ 
    distance  $\leftarrow$  distance + ( $x_{idxFront+1} - x_{idxFront}$ )  $\cdot$ 
    |partialSumFront|
    idxFront  $\leftarrow$  idxFront + 1
while  $x_{idxBack} > 0$  do
    partialSumBack  $\leftarrow$  partialSumBack +  $a_{idxEnd}$ 
    distance  $\leftarrow$  distance + ( $x_{idxBack} - x_{idxBack-1}$ )  $\cdot$ 
    |partialSumBack|
    idxBack  $\leftarrow$  idxBack - 1
for  $idx \leftarrow idxFront$  to idxBack do
    partialSumFront  $\leftarrow$  partialSumFront +  $a_{idx}$ 
return distance + |partialSumFront + partialSumBack|

```

3.2.4 Complexity of the algorithm

Each iteration of each loop takes a constant time. The total number of iterations in all three loops is equal to $k + 1 + (n - k - 1)$. Computational complexity of this algorithm is therefore $\Theta(n)$, while the memory complexity is $\Theta(1)$.

3.3 Computing of flat distance

The algorithm for flat distance requires storing the shape of functions analogous to W_1^m as they get more complicated when m increases. In Section 3.3.2 we provide a recursive formula for the sequence of these functions. The pseudocode in Section 3.3.3 implements the algorithm using an abstract data structure to store previously defined functions. The

computational complexity depends on the choice of this structure. In further sections we provide two solutions that require $O(n^2)$ and $O(n \log n)$ operations.

3.3.1 Reduction to the case of discrete measures

Similarly to the previous cases the following lemma holds.

Lemma 3.3. *Let μ be an arbitrary Radon measure supported on $[0, K]$. For every $\varepsilon > 0$ there exist a discrete measure μ_ε such that*

$$F(\mu, \mu_\varepsilon) < \varepsilon$$

The proof follows from the fact that $\|\mu\| = \|\nu\|$ implies $F(\mu, \nu) \leq W_1(\mu, \nu)$.

3.3.2 The Algorithm

We assume

$$\mu - \nu = \sum_{i=1}^n a_i \delta_{x_i}.$$

Computing of $F(\mu, \nu)$ is equivalent to finding maximum of

$$\left| \sum_{k=1}^n a_k f_k \right|$$

with the following restrictions

$$\begin{aligned} |f_k| &\leq 1, \\ |f_k - f_{k-1}| &\leq |x_k - x_{k-1}|. \end{aligned}$$

Define

$$F^m(x) = \sup \left\{ \left| \sum_{k=1}^m a_k f_k \right| : f_m = x, |f_k - f_{k-1}| \leq |x_k - x_{k-1}|, |f_k| \leq 1 \right\}.$$

Obviously, it holds by definition

$$F(\mu, \nu) = \sup_{x \in [-1, 1]} F^n(x).$$

Observe that

$$\begin{aligned} F^1(x) &= a_1x, \\ F^2(x) &= a_2x + \sup_{f_1 \in [x-d_1, x+d_1] \cap [-1, 1]} F^1(x) = a_2x + \min(|a_1|, a_1x + |a_1|d_1), \end{aligned}$$

$$F^m(x) = a_mx + \sup_{f_{m-1} \in [x-d_{m-1}, x+d_{m-1}] \cap [-1, 1]} F^{m-1}(x).$$

Computing of $F^m(x)$ based on F^{m-1} is more complex than in previous cases, as F^{m-1} is not necessarily monotonic.

Lemma 3.4. *Function F^m is concave for each m .*

Proof. To prove the lemma we will use induction with respect to m . $F^1(x)$ is given as a_1x , so it is indeed concave. Assume F^m is concave. Define

$$F_{max}^{n,d}(x) = \sup_{[x-d, x+d] \cap [-1, 1]} F^n(x).$$

Choose $x, y \in [-1, 1]$. Then, there exist $x' \in B(x, d) \cap [-1, 1]$, $y' \in B(y, d) \cap [-1, 1]$ such that

$$\alpha F_{max}^{m,d}(x) + (1 - \alpha) F_{max}^{m,d}(y) = \alpha F^m(x') + (1 - \alpha) F^m(y').$$

Because F^m is concave, it holds

$$\alpha F^m(x') + (1 - \alpha) F^m(y') \leq F^m(\alpha x' + (1 - \alpha)y') \leq F_{max}^{m,d}(\alpha x + (1 - \alpha)y)$$

The last inequality follows from $\alpha x' + (1 - \alpha)y' \in B(\alpha x + (1 - \alpha)y, d)$. It is now proven that $F^{m+1}(x)$ is concave, as it is a sum of a linear function and a concave function $F_{max}^{m,d}(y)$. \square

Lemma 3.5. *For each m the function F^m is piecewise linear on m intervals and it holds for some point x_m*

$$F^m = a_mx + \begin{cases} F^{m-1}(x + d_{m-1}) & \text{on } [-1, x_m - d_{m-1}] \\ F^{m-1}(x_m) & \text{on } [x_m - d_{m-1}, x_m + d_{m-1}] \\ F^{m-1}(x - d_{m-1}) & \text{on } [x_m + d_{m-1}, 1] \end{cases}$$

Proof. F^1 is a linear function, so it can be described by its values in ± 1 . Assume that F^m can be described by at most $m + 1$ points and is linear between those points. As F^m is concave, there exists a point $x_m \in [-1, 1]$ such that $F^m(x) \leq F^m(x_m)$ for every x . The maximum of F^m on an interval whose both ends are smaller than x_m is taken on its right end. Similarly, if both ends of the intervals are larger than x_m , the maximum is taken on its left end. Finally, if the interval contains x_m , the maximum is in x_m . These considerations prove the formula for F^{m+1} . It follows that F_{m+1} is piecewise linear and it can be described by as many points as $F^n|_{[d, 1-d]}$ plus 1. \square

3.3.3 Pseudocode

In the following code a data structure called 'funcDescription' being a set of pairs will be used to describe F^{idx} . Its has the following interpretation:

1. $F^{idx}(-1) = leftValue$
2. if $(v, p) \in funcDescription$ then $F^{idx}(x)' = p$ for all x larger than v and smaller than the next value in the structure.

The last instruction in the main loop of the pseudocode makes it inefficient to implement 'funcDescription' as a simple BST tree.

Input: Non-decreasing table of positions $x \in [0, 1]^n$, table of masses $a \in \mathbb{R}^n$

FLAT-DISTANCE

```

leftValue  $\leftarrow$  0
funcDescription  $\leftarrow \{(-1, 0), (1, -\infty)\}$ 
for idx  $\leftarrow$  1 to n do
    d  $\leftarrow$   $x_{idx} - x_{idx-1}$ 
    funcLeft  $\leftarrow$ 
     $\{(v - d, p) : (v, p) \in funcDescription \wedge p > 0\}$ 
    funcRight  $\leftarrow$ 
     $\{(v + d, p) : (v, p) \in funcDescription \wedge p < 0\}$ 
     $v_m \leftarrow \min \{v : (v, p) \in funcRight\}$ 
    funcDescription  $\leftarrow funcLeft \cup \{(v_m - 2d, 0)\} \cup funcRight$ 
    leftValue  $\leftarrow$  leftValue +
     $\sum_{(v,p) \in funcDescription, v < 1} (\min(\min \{v' : (v', -) \in funcDescription \wedge v' > v\}, -1) - v) \cdot$ 
     $p$ 
     $(v_{min}, p_{min}) \leftarrow$ 
     $\max \{(v, p) : (v, p) \in funcDescription[i] \wedge v \leq -1\}$ 
     $(v_{max}, p_{max}) \leftarrow$ 
     $\max \{(v, p) : (v, p) \in funcDescription[i] \wedge v \in [0, 1]\}$ 
    funcDescription  $\leftarrow$ 
     $(funcDescription \cap \{(v, p) : v \in [0, 1]\}) \cup$ 
     $\{(\max(v_{min}, -1), p_{min})\} \cup \{(1, -\infty)\}$ 
    funcDescription  $\leftarrow$ 
     $\{(x, p + a_n) : (x, p) \in funcDescription\}$ 
return leftValue +
 $\sum_{(v,p) \in funcDescription, p > 0} (\min \{v' : (v', -) \in funcDescription \wedge v' > v\} - v) \cdot$ 
 $p$ 

```

3.3.4 Complexity $O(n^2)$ of the algorithm

As mentioned before, the complexity of this algorithm depends on the implementation of *funcDescription* data structure.

The simplest implementation of *funcDescription* uses an array of pairs (v, p) sorted by v in ascending order and by p in the reverse order in the same time. This is possible based on Lemma 3.3.2.

The first block of instructions can be performed in $\Theta(\#funcDescription)$ by simply shifting all elements such that $p < 0$ to the right, and modifying v by iterating over all elements of *funcDescription*. The next block (computing of *leftValue*) can be computed with the same complexity, as $\min\{v' : (v', _) \in funcDescription \wedge v' > v\}$ is simply the next element after v in the ordered array. Finally, every instruction in the last block can be performed in $\Theta(\#funcDescription)$ by iterating over all its elements.

In each iteration of the main loop at most 1 element is added to *funcDescription*. Therefore the computational complexity of the algorithm is $O(n^2)$ while the memory complexity is $O(n)$.

3.3.5 Complexity $O(n \log n)$ of the algorithm

The previous result can be improved to $O(n \log n)$ by using balanced binary search trees data structure.

In this implementation *funcDescription* is represented by global variables $p_{modifier}$ and a BST of values $(\Delta v, p)$ where p is the key. An entry (v, p) in this structure is represented as

$$\left(\sum_{(\Delta v', p') \in funcDescription \wedge p' \geq p} \Delta v', p \right)$$

so obtaining an element of *funcDescription* may take linear time.

The advantages of this structure can be easily seen when analyzing the first block of the code. The division of *funcDescription* by the value of p (at first 0) can be achieved in $O(\log n)$. Shifting all elements of those subsets can then be done in a constant time by modifying first elements of those sets. Adding the extra node also requires $O(\log n)$ operations.

Setting *leftValue* may require linear time, but all (apart from one) visited nodes will be removed in the third block. In all iterations of the main loop this instruction takes, therefore, $O(n)$.

Removing nodes with the first coordinate ≤ -1 is obviously done in $O(n)$ in total. Identifying nodes with the first coordinate ≥ 1 might seem problematic. It is, however, known that for the least p the value of v is equal to $1 + d$. Relevant nodes can be, therefore, removed from the back in $O(n)$. Adding a_n to the second coordinate of each node is done by adding it to global variable $p_{modifier}$.

All iterations of the main loop require $O(n \log n + n)$ operations. The memory complexity is also $O(n \log n)$.

3.3.6 Performance of the algorithm for the flat distance

Performance of the algorithm depends on the choice of *funcDescription* data structure. Theoretic bounds for computational complexity are, however, not sufficient to argue about performance of these two options. The first reason is that the each operation in $O(n^2)$ algorithm is much faster than in $O(n \log n)$ in terms of number of instructions. Secondly, hardware architectures provide solutions in which iterating over large tables is accelerated. Finally, the algorithm does reach its theoretical bound only if many points concentrate on a small interval. A gap of size 2 between two points completely cleans *funcDescription* data structure. All that is shown in numerical tests. To measure performance we have used a single core of AMD Athlon II X4 605e processor clocked at 2.3Ghz with 8GB of memory. The results are presented in Figs.1-2.

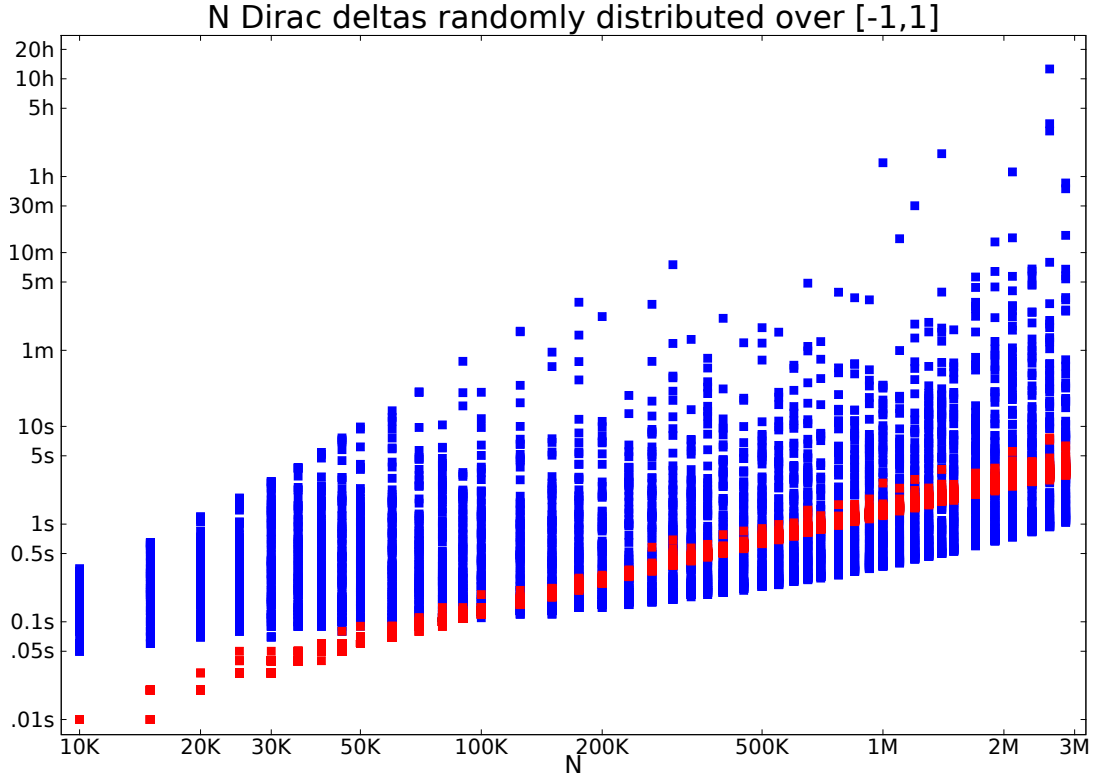


Figure 1: Comparison of the performance of the two proposed algorithms for the flat distance for N Dirac deltas randomly distributed over $[-1, 1]$. The plot shows how the time of computation depends on N . For each input size 100 independent tests were executed to demonstrate how sensitive the algorithms are to input data distribution. Results of $O(n \log n)$ algorithm are depicted as red dots, and results of $O(n^2)$ algorithm as blue dots.

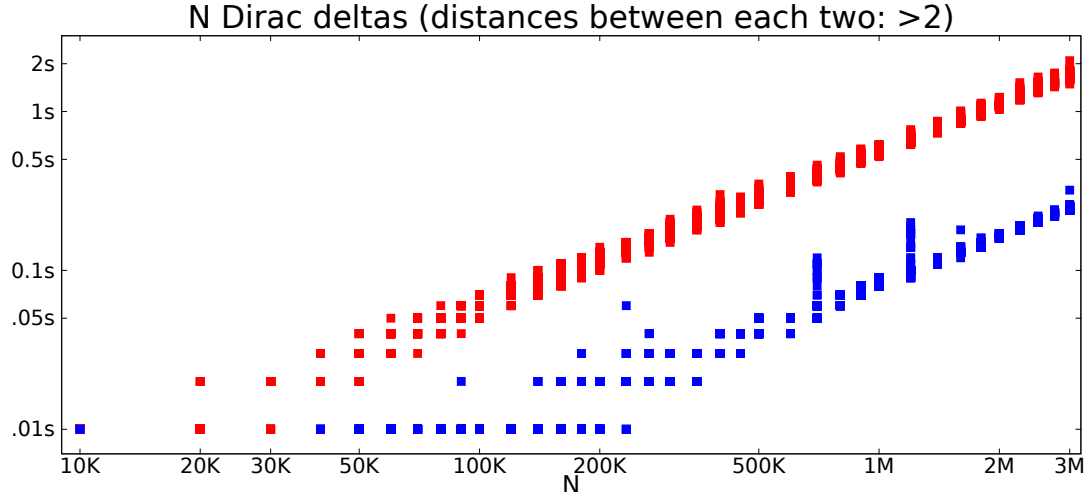


Figure 2: Comparison of the performance of the two proposed algorithms for the flat distance for N Dirac deltas with a distributed over a large domain, i.e. distance between each two masses is larger than 2. In this case both algorithms are in fact linear, as the `funcDescription` structure has at most two elements. The plot demonstrates the overhead of using BST structures. Results of $O(n \log n)$ algorithm are depicted as red dots, and results of $O(n^2)$ algorithm as blue dots.

Acknowledgments

JJ was supported by the International Ph.D. Projects Program of Foundation for Polish Science operated within the Innovative Economy Operational Program 2009-2015 funded by European Regional Development Fund (Ph.D. Program: Mathematical Methods in Natural Sciences). AM-C was supported by the ERC Starting Grant "Biostruct" No. 210680 and the Emmy Noether Program of the German Research Council (DFG).

References

- [1] L. AMBROSIO AND G. CRIPPA. *Existence, Uniqueness, Stability and Differentiability Properties of the Flow Associated to Weakly Differentiable Vector Fields* in Transport Equations and Multi-D Hyperbolic Conservation Laws, Lect. Notes Unione Mat. Ital., Springer 5: 3–57, 2008.
- [2] L. AMBROSIO, N. GIGLI AND G. SAVARÉ. *Gradient Flows in Metric Spaces and in the Space of Probability Measures*, Birkhäuser, ETH Lecture Notes in Mathematics, 2005.
- [3] A. BRANNSTROM, L. CARLSSON, D. SIMPSON. On the convergence of the escalator boxcar train, arXiv:1210.1444v1.
- [4] J.A. CARRILLO, R.M. COLOMBO, P. GWIAZDA, A. ULIKOWSKA. Structured populations, cell growth and measure valued balance laws, *J. Diff. Eq.* 252: 3245–3277, 2012.
- [5] J.A. CARRILLO M. DI FRANCESCO AND G. TOSCANI. Strict contractivity of the 2-Wasserstein distance for the porous medium equation by mass-centering *Proc. Amer. Math. Soc.* 135: 353–363, 2007.
- [6] L.C. EVANS AND R.F. GARIEPY. *Measure Theory and Fine Properties of Functions*, CRC Press, 1992.
- [7] P. GWIAZDA, T. LORENZ AND A. MARCINIAK-CZOCHRA. A nonlinear structured population model: Lipschitz continuity of measure valued solutions with respect to model ingredients, *J. Diff. Eq.* 248: 2703–2735, 2010.
- [8] P. GWIAZDA AND A. MARCINIAK-CZOCHRA, Structured population models in metric spaces, *J. Hyper. Diff. Eq.* 7: 733–773, 2010.
- [9] R. JORDAN, D. KINDERLEHRER AND F. OTTO. The variational formulation of the Fokker-Planck equation *SIAM J. Math. Anal.* 29: 1–17, 1998.
- [10] D. KINDERLEHRER AND N.J. WALKINGTON. Approximation of parabolic equations using the Wasserstein metric. *Math. Mod. Num. Anal.* 33: 837–852, 1999.

- [11] H. NEUNZERT, *An introduction to the nonlinear Boltzmann-Vlasov equation*, in Kinetic Theories and the Boltzmann Equation, Springer, Berlin, Lecture Notes in Math. 1048: 60–110, 1981.
- [12] A. ULIKOWSKA. An age-structured, two-sex model in the space of Radon measures: Well posedness, *Kinetic and Related Models*, 5: 873–900, 2012.
- [13] M. WESTDICKENBERG AND J. WILKENING. Variational particle schemes for the porous medium equation and for the system of isentropic Euler equations *Math. Mod. Num. Anal.* 44: 133–166, 2010.
- [14] C. VILLANI *Optimal transport: old and new*, Springer-Verlag, Berlin, 2009.